Forgetting of the initial distribution for Hidden Markov Models

Randal Douc, Gersende Fort, Éric Moulines, Pierre Priouret

Forgetting of the initial distribution for Hidden Markov Models [1]

R. Douc [a], G. Fort [b,*], E. Moulines [b], P. Priouret [c]

[a] *CMAP, Ecole Polytechnique, 91128 Palaiseau, France.*

[b] *LTCI, CNRS-GET/Télécom Paris, 46 rue Barrault, 75634 Paris Cedex 13, France.*

[c] *LPMA, Univ. P.M. Curie, Boîte courrier 188, 75252 Paris Cedex 05, France.*

Abstract

The forgetting of the initial distribution for discrete Hidden Markov Models (HMM) is addressed: a new set of conditions is proposed, to establish the forgetting property of the filter, at a polynomial and geometric rate. Both a pathwise-type convergence of the total variation distance of the filter started from two different initial distributions, and a convergence in expectation are considered. The results are illustrated using different HMM of interest: the dynamic tobit model, the non-linear state space model and the stochastic volatility model.

Key words: Nonlinear filtering, Hidden Markov Models, asymptotic stability, total variation norm.

1991 MSC: 93E11, 60B10, 60G35

1 Definition and notations

A Hidden Markov Model (HMM) is a doubly stochastic process with an underlying Markov chain that is not directly observable. More specifically, let X and Y be two spaces equipped with a countably generated σ-fields $\mathcal{X}$ and $\mathcal{Y}$; denote by Q and G respectively, a Markov transition kernel on $(\mathsf{X}, \mathcal{X})$ and a transition kernel from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Y}, \mathcal{Y})$. Consider the Markov transition kernel defined for any $(x, y) \in \mathsf{X} \times \mathsf{Y}$ and $C \in \mathcal{X} \otimes \mathcal{Y}$ by

$$T\left[(x, y), C\right] \stackrel{\text{def}}{=} Q \otimes G[(x, y), C] = \iint Q(x, dx')\, G(x', dy') \mathbb{1}_C(x', y') \,. \tag{1}$$

We consider $\{X_k, Y_k\}_{k \geq 0}$ the Markov chain with transition kernel T and initial distribution $\nu \otimes G(C) \stackrel{\text{def}}{=} \iint \nu(dx) G(x, dy) \mathbb{1}_C(x, y)$, where ν is a probability measure on $(\mathsf{X}, \mathcal{X})$. We assume that the chain $\{X_k\}_{k \geq 0}$ is not observable (hence the name *hidden*). The model is said to be partially dominated if there exists a measure μ on $(\mathsf{Y}, \mathcal{Y})$ such that for all $x \in \mathsf{X}$, $G(x, \cdot)$ is absolutely continuous with respect to μ: in such case, the joint transition kernel T can be written as

$$T\left[(x, y), C\right] = \iint Q(x, dx') g(x', y')\, \mathbb{1}_C(x', y') \mu(dy') \,, \quad C \in \mathcal{X} \otimes \mathcal{Y} \,, \tag{2}$$

where $g(x, \cdot) = \frac{dG(x, \cdot)}{d\mu}$ denotes the Radon-Nikodym derivative of $G(x, \cdot)$ with respect to μ. To follow the usage in the filtering literature, $g(x, \cdot)$ is referred to as the *likelihood* of the observation. An example of such type of dependence is $X_{k+1} = a(X_k, \zeta_{k+1})$ and $Y_k = b(X_k, \varepsilon_k)$, where $\{\zeta_k\}_{k \geq 0}$ and $\{\varepsilon_k\}_{k \geq 0}$ are i.i.d. sequences of random variables, and $\{\zeta_k\}_{k \geq 0}$, $\{\varepsilon_k\}_{k \geq 0}$ and X_0 are independent. The most elementary example is the so-called linear Gaussian state space model (LGSSM) where a and b are linear and $\{\zeta_k, \varepsilon_k\}_{k \geq 0}$ are

* Corresponding author.
Email addresses: douc@cmapx.polytechnique.fr (R. Douc), gfort@tsi.enst.fr (G. Fort), moulines@tsi.enst.fr (E. Moulines), priouret@ccr.jussieu.fr (P. Priouret).
[1] This work was partly supported by the National Research Agency (ANR) under the program "ANR-05-BLAN-0299".

i.i.d. standard Gaussian. We denote by $\phi_{\nu,n}[y_{0:n}]$ the distribution of the hidden state X_n conditionally on the observations $y_{0:n} \stackrel{\text{def}}{=} [y_0, \ldots, y_n]$, which is given by

$$\begin{aligned}\phi_{\nu,n}[y_{0:n}](A) \stackrel{\text{def}}{=} & \frac{\nu\left[g(\cdot, y_0) Q g(\cdot, y_1) Q \ldots Q g(\cdot, y_n) \mathbb{1}_A\right]}{\nu\left[g(\cdot, y_0) Q g(\cdot, y_1) Q \ldots Q g(\cdot, y_n)\right]} \\ & = \frac{\int_{\mathsf{X}^{n+1}} \nu(dx_0) g(x_0, y_1) \prod_{i=1}^n Q(x_{i-1}, dx_i) g(x_i, y_i) \mathbb{1}_A(x_n)}{\int_{\mathsf{X}^{n+1}} \nu(dx_0) g(x_0, y_1) \prod_{i=1}^n Q(x_{i-1}, dx_i) g(x_i, y_i)} \ , \end{aligned} \quad (3)$$

where $Qf(x) = Q(x, f) \stackrel{\text{def}}{=} \int Q(x, dx') f(x')$, for any function $f \in \mathbb{B}_+(\mathsf{X})$ the set of non-negative functions $f : \mathsf{X} \to \mathbb{R}$, such that f is $\mathcal{X}/\mathcal{B}(\mathbb{R})$ measurable, with $\mathcal{B}(\mathbb{R})$ the Borel σ-algebra.

In practice the model is rarely known exactly and so suboptimal filters are constructed by replacing the unknown transition kernel, likelihood function and initial distribution by suitable approximations.

The choice of these quantities plays a key role both when studying the convergence of sequential Monte Carlo methods or when analysing the asymptotic behaviour of the maximum likelihood estimator (see e.g. (8) or (5) and the references therein).

The simplest problem assumes that the transitions are known, so that the only error in the filter is due to a wrong initial condition. A typical question is to ask whether $\phi_{\nu,n}[y_{0:n}]$ and $\phi_{\nu',n}[y_{0:n}]$ are close (in some sense) for large values of n, and two different choices of the initial distribution ν and ν'.

The forgetting property of the initial condition of the optimal filter in nonlinear state space models has attracted many research efforts and it would be a formidable task to give credit to every contributors. The purpose of the short presentation of the existing results below is mainly to allow comparison of assumptions and results presented in this contributions with respect to those previously reported in the literature. The first result in this direction has been obtained by (21), who established L_p-type convergence of the optimal filter initialised with the wrong initial condition to the filter initialised with the true initial distribution (assuming that the transition kernels are known); however, their proof does not provide a rate of convergence. A new approach based on the Hilbert projective metric has later been introduced in (2) to obtain the exponential stability of the optimal filter with respect to its initial condition. However their results were based on stringent *mixing* conditions for the transition kernels; these conditions state that there exist positive constants ϵ_- and ϵ_+ and a probability measure λ on $(\mathsf{X}, \mathcal{X})$ such that for $f \in \mathbb{B}_+(\mathsf{X})$,

$$\epsilon_- \lambda(f) \leq Q(x, f) \leq \epsilon_+ \lambda(f) \;, \quad \text{for any } x \in \mathsf{X} \;. \tag{4}$$

This condition in particular implies that the chain is uniformly geometrically ergodic. Similar results were obtained independently by (9) using the Dobrushin ergodicity coefficient (see (11) for further refinements under this assumption). The mixing condition has later been weakened by (6), under the assumption that the kernel Q is positive recurrent and is dominated by some reference measure λ:

$$\sup_{(x,x') \in \mathsf{X} \times \mathsf{X}} q(x, x') < \infty \quad \text{and} \quad \int \operatorname{essinf} q(x, x') \pi(x) \lambda(dx) > 0 \;,$$

where $q(x, \cdot) = \frac{dQ(x,\cdot)}{d\lambda}$, essinf is the essential infimum with respect to λ and $\pi d\lambda$ is the stationary distribution of the chain Q . If the upper bound is reasonable, the lower bound is restrictive in many applications and fails to be satisfied e.g. for the linear state space Gaussian model.

In (18), the stability of the optimal filter is studied for a class of kernels referred to as *pseudo-mixing*. The definition of pseudo-mixing kernel is adapted to the case where the state space is $\mathsf{X} = \mathbb{R}^d$, equipped with the Borel sigma-field $\mathcal{X}$. A kernel Q on $(\mathsf{X}, \mathcal{X})$ is *pseudo-mixing* if for any compact set C with a diameter d large enough, there exist positive constants $\epsilon_-(d) > 0$ and $\epsilon_+(d) > 0$ and a measure λ_{C} (which may be chosen to be finite without loss of generality) such that

$$\epsilon_-(d)\lambda_{\mathsf{C}}(A) \leq Q(x, A) \leq \epsilon_+(d)\lambda_{\mathsf{C}}(A) , \quad \text{for any } x \in \mathsf{C},\ A \in \mathcal{X} \tag{5}$$

This condition implies that for any $(x', x'') \in \mathsf{C} \times \mathsf{C}$,

$$\frac{\epsilon_-(d)}{\epsilon_+(d)} < \operatorname{essinf}_{x \in \mathsf{X}} q(x', x)/q(x'', x) \leq \operatorname{esssup}_{x \in \mathsf{X}} q(x', x)/q(x'', x) \leq \frac{\epsilon_+(d)}{\epsilon_-(d)} ,$$

where $q(x, \cdot) \stackrel{\text{def}}{=} dQ(x, \cdot)/d\lambda_{\mathsf{C}}$, and esssup and essinf denote the essential supremum and infimum with respect to λ_{C}. This condition is obviously more general than (4), but still it is not satisfied in the linear Gaussian case (see (18, Example 4.3)).

Several attempts have been made to establish the stability conditions under the so-called *small* noise condition. The first result in this direction has been obtained by (2) (in continuous time) who considered an ergodic diffusion process with constant diffusion coefficient and linear observations: when the variance of the observation noise is sufficiently small, (2) established that the filter is exponentially stable. Small noise conditions also appeared (in a discrete time setting) in (4) and (22). These results do not allow to consider the linear Gaussian state space model with arbitrary noise variance.

A very significant step has been achieved by (16), who considered the filtering problem of Markov chain $\{X_k\}_{k\geq 0}$ with values in $\mathsf{X} = \mathbb{R}^d$ filtered from observations $\{Y_k\}_{k\geq 0}$ in $\mathsf{Y} = \mathbb{R}^\ell$,

$$\begin{cases} X_{k+1} = X_k + b(X_k) + \sigma(X_k)\zeta_k , \\ Y_k = h(X_k) + \beta\varepsilon_k . \end{cases} \tag{6}$$

Here $\{(\zeta_k, \varepsilon_k)\}_{k\geq 0}$ is a i.i.d. sequence of standard Gaussian random vectors in $\mathbb{R}^{d+\ell}$, $b(\cdot)$ is a d-dimensional vector function, $\sigma(\cdot)$ a $d \times d$-matrix function, $h(\cdot)$ is a ℓ-dimensional vector-function and $\beta > 0$. The author established, under appropriate conditions on b, h and σ, that the optimal filter forgets the initial conditions; these conditions cover (with some restrictions) the linear Gaussian state space model.

In this contribution, we will propose a new set of conditions to establish the forgetting property of the filter, which are more general than those proposed in (16). In theorem 1, a pathwise-type convergence of the total variation distance of the filter started from two different initial distributions is established, which is shown to hold almost surely w.r.t. the probability distribution of the observation process $\{Y_k\}_k$. Then, in Theorem 3, the convergence of the expectation of this total variation distance is shown, under more stringent conditions. The results are shown to hold under rather weak conditions on the observation process $\{Y_k\}_k$ which do not necessarily entail that the observations are from an HMM.

The paper is organised as followed. In section 2, we introduce the assumptions and state the main results. In section 3, we give sufficient conditions for Theorems 1 and 3 to hold, when $\{Y_k\}_k$ is an HMM process, assuming that the transition kernel and the likelihood function might be different from those used in the definition of the filter. In section 4, we illustrate the use of our assumptions on several examples with unbounded state spaces. The proofs are given in sections 5 and 6.

2 Assumptions and Main results

We say that a set $\mathsf{C} \in \mathcal{X}$ satisfies the *local Doeblin* property (for short, C is a LD-set), if there exists a measure λ_{C} and constants $\epsilon_{\mathsf{C}}^{-} > 0$ and $\epsilon_{\mathsf{C}}^{+} > 0$ such that, $\lambda_{\mathsf{C}}(\mathsf{C}) > 0$ and for any $A \in \mathcal{X}$,

$$\epsilon_{\mathsf{C}}^{-}\lambda_{\mathsf{C}}(A \cap \mathsf{C}) \leq Q(x, A \cap \mathsf{C}) \leq \epsilon_{\mathsf{C}}^{+}\lambda_{\mathsf{C}}(A \cap \mathsf{C}) , \quad \text{for all } x \in \mathsf{C} . \tag{7}$$

Locally Doeblin sets share some similarities with 1-small set in the theory of Markov chains over general state spaces (see (20, chapter 5)). Recall that a set C is 1-small if there exists a measure $\tilde{\lambda}_{\mathsf{C}}$ and $\tilde{\epsilon}_{\mathsf{C}} > 0$, such that $\tilde{\lambda}_{\mathsf{C}}(\mathsf{C}) > 0$, and for all $x \in \mathsf{C}$ and $A \in \mathcal{X}$,

$Q(x, A \cap \mathsf{C}) \geq \tilde{\epsilon}_{\mathsf{C}} \tilde{\lambda}_{\mathsf{C}}(A \cap \mathsf{C})$. In particular, a locally Doeblin set is 1-small with $\tilde{\epsilon}_{\mathsf{C}} = \epsilon^{-}_{\mathsf{C}}$ and $\tilde{\lambda}_{\mathsf{C}} = \lambda_{\mathsf{C}}$. The main difference stems from the fact that we impose both a lower and an upper bound, and we impose that the minorizing and the majorizing measure are the same.

Compared to the pseudo-mixing condition (5), the local Doeblin property involves the trace of the Markov kernel Q on C and thus happens to be much less restrictive. In particular, on the contrary to the pseudo-mixing condition, it can be easily checked that for the kernel associated to the linear Gaussian state space model, every bounded Borel set C is locally Doeblin.

Let V be a positive function $V : \mathsf{X} \to [1, \infty)$ and $A \in \mathcal{X}$ be a set. Define:

$$\Upsilon_{\mathsf{A}}(y) \stackrel{\text{def}}{=} \sup_{x \in \mathsf{A}} g(x, y) QV(x)/V(x) \; . \tag{8}$$

Consider the following assumptions:

(H1) For any $(x, y) \in \mathsf{X} \times \mathsf{Y}$, $g(x, y) > 0$.

(H2) There exist a set $\mathsf{K} \subseteq \mathsf{Y}$ and a function $V : \mathsf{X} \to [1, \infty)$ such that for any $\eta > 0$, one may choose a LD-set $\mathsf{C} \in \mathcal{X}$ satisfying

$$\Upsilon_{\mathsf{C}^c}(y) \leq \eta \, \Upsilon_{\mathsf{X}}(y) \; , \quad \text{for all } y \in \mathsf{K}.$$

Assumption (H1) can be relaxed, but this assumption simplifies the statements of the results and the proofs. The case where the likelihood may vanish will be considered in a companion paper. Assumption (H2) involves both the likelihood function and the drift function. It is satisfied for example if there exists a set K such that for all $\eta > 0$, one can choose a LD-set C so that

$$\sup_{x \in \mathsf{C}^c} g(x, y) < \eta \sup_{x \in \mathsf{X}} g(x, y) \; , \quad \text{for all } y \in \mathsf{K}, \tag{9}$$

in which case the previous assumption is satisfied with $V \equiv 1$. When $\mathsf{X} = \mathbb{R}^d$, this situation occurs for example when the compact sets are locally Doeblin and $\lim_{|x|\to\infty} \sup_{y\in\mathsf{K}} g(x,y) = 0$. As a simple illustration, this last property is satisfied for $Y_k = h(X_k) + \epsilon_k$ with $\lim_{|x|\to\infty} |h(x)| = \infty$ and $\{\epsilon_k\}_k$ are i.i.d.random variables (independent of $\{X_k\}_k$) with a density g which satisfies $\lim_{|x|\to\infty} g(x) = 0$. More complex models satisfying (H2) are considered in Section 4.

When (9) is not satisfied, assumption (H2) can still be fulfilled if for all $y \in \mathsf{Y}$, $\sup_{x\in\mathsf{X}} g(x,y) < \infty$, $\sup_{\mathsf{X}} QV/V < \infty$ for some function $V : \mathsf{X} \to [1,\infty)$, and for all $\eta > 0$, there exists a LD-set C such that $\sup_{\mathsf{C}^c} QV/V \leq \eta$. As a simple illustration, this situation occurs for example with $X_{k+1} = \phi X_k + \sigma\zeta_k$ where $|\phi| < 1$, $\sigma > 0$ and $\{\zeta_k\}_k$ a family of iid standard Gaussian vectors. More details are provided in Section 4.

For any LD-set D and ν a probability measure on $(\mathsf{X}, \mathcal{X})$ define:

$$\Phi_{\nu,\mathsf{D}}(y_0, y_1) \stackrel{\text{def}}{=} \nu[g(\cdot, y_0) Q g(\cdot, y_1) \mathbb{1}_{\mathsf{D}}] \;, \tag{10}$$

$$\Psi_{\mathsf{D}}(y) \stackrel{\text{def}}{=} \lambda_{\mathsf{D}}(g(\cdot, y)\mathbb{1}_{\mathsf{D}}) \;. \tag{11}$$

We denote by $(\Omega, \mathcal{A})$ a measurable space, and we let $\{Y_k\}_{k\geq 0}$ be a stochastic process on $(\Omega, \mathcal{A})$ which takes values in $(\mathsf{Y}, \mathcal{Y})$ but *which is not necessarily the observation of an HMM*. For any probability measure ν and any $n \in \mathbb{N}$, the filtering distribution $\phi_{\nu,n}[Y_{0:n}]$ (defined in (3)) is a measure-valued random variable on $(\Omega, \mathcal{A})$.

Theorem 1 *Assume (H1-2) and let $\mathbb{P}_\star$ be a probability measure on $(\Omega, \mathcal{A})$. Assume in addition that for some LD-set D and some constants $M > 0$ and $\gamma \in (0,1)$,*

$$\liminf_{n\to\infty} n^{-1} \sum_{i=0}^{n} \mathbb{1}_{\mathsf{K}}(Y_i) \geq (1+\gamma)/2 \;, \qquad \mathbb{P}_\star - a.s. \tag{12}$$

$$\limsup_{n\to\infty} n^{-1} \sum_{i=0}^{n} \log \Upsilon_{\mathsf{X}}(Y_i) < M \;, \qquad \mathbb{P}_\star - a.s. \tag{13}$$

$$\liminf_{n\to\infty} n^{-1} \sum_{i=2}^{n} \log \Psi_{\mathsf{D}}(Y_i) > -M \;, \qquad \mathbb{P}_\star - a.s. \tag{14}$$

where Υ_{X} and Ψ_{D} are defined in (8) *and* (11)*, respectively. Then, for any initial distributions ν and ν' on $(\mathsf{X}, \mathcal{X})$ such that $\nu(V) + \nu'(V) < \infty$, $\nu Q \mathbb{1}_{\mathsf{D}} > 0$ and $\nu' Q \mathbb{1}_{\mathsf{D}} > 0$, there exists a positive constant c such that,*

$$\limsup_{n\to\infty} n^{-1} \log \|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}} < -c \ , \quad \mathbb{P}_\star - a.s. \tag{15}$$

Remark 2 *We stress that it is not necessary to assume that $\{Y_k\}_{k\geq 0}$ is the observation of an HMM $\{X_k, Y_k\}_{k\geq 0}$. Conditions* (13) *and* (14) *can be verified for example under a variety of weak dependence conditions, the only requirement being basically to be able to prove a LLN (see for example (7)). This is of interest because in many applications, the HMM model is not correctly specified, but it is still of interest to establish the forgetting properties of the filtering distribution with respect to the initial distribution.*

We will now state a statement allowing to control the expectation of the total variation distance.

Theorem 3 *Assume (H2). Let D be a LD-set. Then, for any $M_i > 0$, $i = 0, 1, 2$, and $\gamma \in (0, 1)$, there exist $\beta \in (0, 1)$ such that, for any given initial distributions ν and ν' on $(\mathsf{X}, \mathcal{X})$ and all n,*

$$\mathbb{E}_\star \left(\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}} \right) \\ \leq \beta^n \ [1 + \nu(V)\nu'(V)] + r_0(\nu, n) + r_0(\nu', n) + \sum_{i=1}^{3} r_i(n) \tag{16}$$

where the sequences $\{r_0(\nu, n)\}_{n\geq 0}$ and $\{r_i(n)\}_{n\geq 0}$, $i = 1, 2, 3$ are defined by

$$r_0(\nu, n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(\log \Phi_{\nu,\mathsf{D}}(Y_0, Y_1) \leq -M_0 n\right) \ , \tag{17}$$

$$r_1(n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(\sum_{i=0}^{n} \log \Upsilon_{\mathsf{X}}(Y_i) \geq M_1 n \right) \ , \tag{18}$$

$$r_2(n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(\sum_{i=0}^{n} \log \Psi_{\mathsf{D}}(Y_i) \leq -M_2 n \right) \ , \tag{19}$$

$$r_3(n) \stackrel{\text{def}}{=} \mathbb{P}_\star \left(n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\mathsf{K}}(Y_i) \leq (1 + \gamma)/2 \right) \ . \tag{20}$$

3 Applications to HMM

We will now discuss conditions upon which (13) and (14) hold (Propositions 4 to 6) and upon which the right hand side in (16) vanishes (Proposition 7 to Corollary 11). To that goal, we assume that $\{Y_k\}_{k\geq 0}$ is the observation of an HMM $\{X_k, Y_k\}_{k\geq 0}$ with Markov kernel $T_\star = Q_\star \otimes G_\star$, where $Q_\star$ is a transition kernel on $(\mathsf{X}, \mathcal{X})$ and $G_\star$ is a Markov kernel from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Y}, \mathcal{Y})$, and initial distribution $\nu_\star$ on $(\mathsf{X}, \mathcal{X})$.

Recall that a kernel P on a general state space $(\mathsf{Z}, \mathcal{Z})$ is phi-irreducible and (strongly) aperiodic if there exists a σ-finite measure φ on $(\mathsf{Z}, \mathcal{Z})$, such that, for any $A \in \mathcal{Z}$ satisfying $\varphi(A) > 0$ and any initial condition x, $P^n(x; A) > 0$, for all n sufficiently large. A set $\mathsf{C} \in \mathcal{Z}$ is called *petite* for the Markov kernel P if for some probability measure m on $\mathbb{N}$, with finite mean sampling time (which can always be done without loss of generality (20, Proposition 5.5.6))

$$\sum_{n=0}^{\infty} m(n) P^n(x, A) \geq \epsilon_{\mathsf{C}}^- \lambda_{\mathsf{C}}(A) \ , \quad \text{for all } x \in \mathsf{C},\ A \in \mathcal{Z},$$

where λ_{C} is a measure on $(\mathsf{Z}, \mathcal{Z})$ satisfying $\lambda_{\mathsf{C}}(\mathsf{C}) > 0$ and $\epsilon_{\mathsf{C}}^- > 0$. We denote by $\mathbb{P}_\nu^P$ and $\mathbb{E}_\nu^P$ the probability distribution and the expectation on the canonical probability space $(\mathsf{Z}^{\mathbb{N}}, \mathcal{Z}^{\otimes \mathbb{N}})$ associated to the Markov chain with transition kernel P and initial distribution ν.

We first state sufficient conditions for $T_\star$ to be an aperiodic positive Harris chain (see definitions and main properties in (20, Chapters 10 & 13) and (5, Chapter 14)) and for the law of large numbers to hold for the Markov chain with kernel $T_\star$.

Proposition 4 *Assume that $Q_\star$ is an aperiodic, positive Harris Markov kernel with stationary distribution $\pi_\star$. Then, the kernel $T_\star$ defined by*

$$T_\star[(x, y), A] \stackrel{\text{def}}{=} \iint Q_\star(x, dx') G_\star(x', dy') \mathbb{1}_A(x', y') \ , \quad A \in \mathcal{X} \otimes \mathcal{Y} \ ,$$

is an aperiodic positive Harris Markov kernel with stationary distribution $\pi_\star \otimes G_\star$. In addition, for any initial distribution $\nu_\star$ on $(\mathsf{X}, \mathcal{X})$, and any function $\varphi \in \mathbb{B}_+(\mathsf{X} \times \mathsf{Y})$ satisfying $\pi_\star \otimes G_\star(\varphi) < \infty$,

$$n^{-1} \sum_{i=0}^{n} \varphi(X_i, Y_i) \to \pi_\star \otimes G_\star(\varphi) \quad \mathbb{P}^{T_\star}_{\nu_\star \otimes G_\star} - a.s. \tag{21}$$

Corollary 5 *If $\pi_\star \otimes G_\star \left(\log \Upsilon_{\mathsf{X}}\right)_+ < \infty$ (resp. $\pi_\star \otimes G_\star \left(\log \Psi_{\mathsf{D}}\right)_- < \infty$), then, condition* (13) *(resp.* (14)*) is satisfied with $\mathbb{P}_\star \stackrel{\text{def}}{=} \mathbb{P}^{T_\star}_{\nu_\star \otimes G_\star}$.*

In many problems of interest, it is not straightforward to establish that the chain is positive Harris; in addition, the distribution $\pi_\star$ is not known explicitly making the conditions of Corollary 5 difficult to check. It is often interesting to apply the following result which is a direct consequence of the f-norm ergodic theorem and the law of large numbers for positive Harris chain (see for example (20, Theorems 14.0.1, 17.0.1)).

Proposition 6 *Let $f_\star \geq 1$ be a function on X. Assume that $Q_\star$ is a phi-irreducible Markov kernel and that there exist a petite set $\mathsf{C}_\star$, a function $V_\star : \mathsf{X} \to [1, \infty)$, and a constant $b_\star$ satisfying*

$$Q_\star V_\star(x) \leq V_\star(x) - f_\star(x) + b_\star \mathbb{1}_{\mathsf{C}_\star}(x) \; . \tag{22}$$

Then, the kernel $Q_\star$ is positive Harris with invariant probability $\pi_\star$ and $\pi_\star(f_\star) < +\infty$. Let $\varphi \in \mathbb{B}_+(\mathsf{X} \times \mathsf{Y})$ be a function such that

$$\sup_{x \in \mathsf{X}} f_\star^{-1}(x) G_\star\left(x, \varphi(x, \cdot)\right) < \infty, \tag{23}$$

Then, $\pi_\star \otimes G_\star(\varphi) < \infty$.

We now derive conditions to compute a bound for $\{r_0(\nu, n)\}_{n \geq 0}$.

Proposition 7 *Assume (H1-2) and that the drift function V defined in (H2) satisfies $\sup_{\mathsf{X}} V^{-1} Q V < \infty$.*

(i) If for some $p \geq 1$,

$$\sup_{i=0,1} \sup_{\mathsf{X}} V^{-1} \mathbb{E}_\star [\log g(\cdot, Y_i)]_-^p < \infty \;, \tag{24}$$

then, there exists a constant C such that, for any initial probability measure ν on $(\mathsf{X}, \mathcal{X})$ such that $\nu Q \mathbb{1}_D > 0$ and all $n \geq 0$, $r_0(\nu, n) \leq C n^{-p} \nu(V)$.

(ii) If for some positive λ,

$$\sup_{i=0,1} \sup_{\mathsf{X}} V^{-1} \mathbb{E}_\star \left(\exp(\lambda [\log g(\cdot, Y_i)]_-) \right) < \infty \;, \tag{25}$$

then there exist positive constants $C, \delta > 0$, such that for any initial probability measure ν on $(\mathsf{X}, \mathcal{X})$ such that $\nu Q \mathbb{1}_D > 0$, and all $n \geq 0$, $r_0(\nu, n) \leq C \mathrm{e}^{-\delta n} \nu(V)$.

To determine the rate of convergence of the sequences $\{r_i(n)\}_{n \geq 0}$ to zero, $i = 1, 2, 3$, it is required to use deviation inequalities for partial sums of the observations $\{Y_k\}_{k \geq 0}$. There are a variety of techniques to prove such results, depending on the type of assumptions which are available. If polynomial rates are enough, then one can apply the standard Markov inequality together with the Marcinkiewicz-Siegmund inequality; see for example (7) or (12).

Proposition 8 *Assume that*

(i) $Q_\star$ is aperiodic and positive Harris Markov kernel with stationary distribution $\pi_\star$.

(ii) There exist a petite set $\mathsf{C}_\star$ and functions $U_\star, V_\star, W_\star : \mathsf{X} \to [1, \infty)$ and a constant $b_\star$ satisfying $\pi_\star(W_\star) < \infty$ and

$$Q_\star V_\star \leq V_\star - U_\star + b_\star \mathbb{1}_{\mathsf{C}_\star} \;,$$

$$Q_\star W_\star \leq W_\star - V_\star + b_\star \mathbb{1}_{\mathsf{C}_\star}$$

Let $p \geq 1$. There exists a constant $C < \infty$ such that for any function φ on $(\mathsf{Y}, \mathcal{Y})$ satisfying $\sup_{\mathsf{X}} U_\star^{-1} G_\star(\cdot, |\varphi|^p) < \infty$ and $\sup_{\mathsf{X}} U_\star^{-1} V_\star^{1-1/p}\ G_\star(\cdot, |\varphi|) < \infty$, and for any *initial probability distribution* $\nu_\star$ *on* $(\mathsf{X}, \mathcal{X})$*, and any* $\delta > 0$*,*

$$\mathbb{P}^{T_\star}_{\nu_\star \otimes G_\star}\left[\sum_{i=1}^{n} \{\varphi(Y_i) - \pi_\star \otimes G_\star(\varphi)\} \geq \delta n\right] \leq C\delta^{-p} n^{-(p/2 \vee 1)} \nu_\star(W_\star) \ ,$$

Corollary 9 *If there exists* $p \geq 1$ *such that*

$$\sup_{\mathsf{X}} f_\star^{-1} G_\star(\cdot, |\log \Upsilon_{\mathsf{X}}|^p) < \infty \ , \qquad \sup_{\mathsf{X}} f_\star^{-1} V_\star^{1-1/p} G_\star(\cdot, |\log \Upsilon_{\mathsf{X}}|) < \infty \ ,$$

and

$$\sup_{\mathsf{X}} f_\star^{-1} G_\star(\cdot, |\log \Psi_{\mathsf{D}}|^p) < \infty \ , \qquad \sup_{\mathsf{X}} f_\star^{-1} V_\star^{1-1/p} G_\star(\cdot, |\log \Psi_{\mathsf{D}}|) < \infty \ ,$$

then there exist finite constants C, M_i*,* $i = 1, 2, 3$ *such that*

$$r_i(n) \leq C n^{-(p/2 \vee 1)} \nu_\star(W_\star) \ .$$

If we wish to establish that the sequences $\{r_i(n)\}_{n \geq 0}$ decreases to zero exponentially fast, we might for example use the multiplicative ergodic theorem (17, Theorem 1.2) to bound an exponential moment of the partial sum, and then use the Markov inequality. This will require to check the multiplicative analog of the additive drift condition (22).

Some additional definitions are needed. Let $W : \mathsf{X} \to (0, \infty)$ be a function. We say that the function W is *unbounded* if $\sup_{\mathsf{X}} W = +\infty$. We define by $\mathcal{G}_W$ the set of functions whose growth at infinity is lower than W, *i.e.* F belongs to $\mathcal{G}_W$ if and only if

$$\sup_{\mathsf{X}} (|F| - W) < \infty \ . \tag{26}$$

Proposition 10 *Let* $W_\star$ *be an unbounded function* $W_\star : \mathsf{X} \to (0, \infty)$ *and that the level sets* $\{W_\star \leq r\}$ *are petite. Assume that* $Q_\star$ *is phi-irreducible and that there exist a function* $V_\star : \mathsf{X} \to [1, \infty)$*, and constant* $b_\star < \infty$ *such that*

$$\log \left(V_\star^{-1} Q_\star V_\star\right) \leq -W_\star + b_\star \ . \tag{27}$$

Then, $Q_\star$ is positive Harris with a unique invariant probability distribution $\pi_\star$, satisfying $\pi_\star(V_\star) < \infty$. Let φ be a non-negative function. If for some $\lambda_\star > 0$,

$$\log\left[G_\star\left(\cdot, \mathrm{e}^{\lambda_\star \varphi}\right)\right] \in \mathcal{G}_{W_\star} \;, \tag{28}$$

there exists a constant $M > 0$ such that, for any initial distribution $\nu_\star$ satisfying $\nu_\star(V_\star) < \infty$,

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}^{T_\star}_{\nu_\star \otimes G_\star}\left(\sum_{i=0}^{n} \varphi(Y_i) \geq Mn\right) < 0 \;. \tag{29}$$

Corollary 11 *Assume that for some $\lambda_\star > 0$,*

$$\log\left[G_\star\left(\cdot, \mathrm{e}^{\lambda_\star [\log \Upsilon_{\mathsf{X}}]_+}\right)\right] \in \mathcal{G}_{W_\star} \qquad \log\left[G_\star\left(\cdot, \mathrm{e}^{\lambda_\star [\log \Psi_{\mathsf{D}}]_-}\right)\right] \in \mathcal{G}_{W_\star} \;.$$

Then, there exist constants M_i, $i = 1, 2$ such that $\limsup_{n\to\infty} n^{-1} \log r_i(n) < 0$, where $\{r_i(n)\}_{n\geq 0}$ are defined in (18) *and* (19)*.*

4 Examples

In this section, we illustrate our results using different models of interest.

4.1 The dynamic tobit model

The tobit model is simply the time series extension of the standard univariate tobit model and so the univariate hidden process is only observed when it is positive ((19) and (1)):

$$\begin{cases} X_{k+1} = \phi X_k + \sigma \zeta_k \;, \\ Y_k = \max(X_k + \beta \varepsilon_k, 0) \;, \end{cases} \tag{30}$$

where $\{(\zeta_k, \varepsilon_k)\}_{k\geq 0}$ is a sequence of i.i.d. standard Gaussian vectors, and $|\phi| < 1$, $\sigma > 0$ and $\beta > 0$. Here $\mathsf{X} = \mathbb{R}$, $\mathsf{Y} = \mathbb{R}_+$ and $\mathcal{X}$ and $\mathcal{Y}$ are the corresponding Borel σ-algebra. The model is partially dominated (see (2)) with respect to the dominating measure $\delta_0 + \lambda^{\mathrm{Leb}}$, where λ^{Leb} is the Lebesgue measure and δ_0 is the Dirac mass at zero. The transition kernels $Q_{\phi,\sigma}$ and the likelihood g_β are respectively given by:

$$Q_{\phi,\sigma}(x, A) = \left(2\pi\sigma^2\right)^{-1/2} \int \exp\left[-(1/2\sigma^2)(x' - \phi x)^2\right] \mathbb{1}_A(x')\lambda^{\mathrm{Leb}}(dx') \,, \tag{31}$$

$$\begin{aligned} g_\beta(x, y) = \mathbb{1}\{y = 0\}(2\pi\beta^2)^{-1/2} \int_x^\infty \exp\left[-(1/2\beta^2)v^2\right] \lambda^{\mathrm{Leb}}(dv) \\ + \mathbb{1}\{y > 0\}(2\pi\beta^2)^{-1/2} \exp\left[-(1/2\beta^2)(y - x)^2\right] \,. \end{aligned} \tag{32}$$

We denote $Q = Q_{\phi,\sigma}$ and $g = g_\beta$.

We assume that $\{Y_k\}_{k\geq 0}$ are the observations of a tobit model (30) with initial distribution $\nu_\star$ and 'parameters' $\phi_\star$, $\sigma_\star$, $\beta_\star$ (which may be different from ϕ, σ, β) satisfying $|\phi_\star| < 1$, $\sigma_\star > 0$ and $\beta_\star > 0$. We denote by $Q_\star = Q_{\phi_\star,\sigma_\star}$, $G_\star(x, \cdot) = g_{\beta_\star}(x, \cdot)\lambda^{\mathrm{Leb}}$ and $\mathbb{E}_\star = \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star}$, where $T_\star = Q_\star \otimes G_\star$.

4.1.1 Assumptions H1 and H2

It is easily seen that any bounded Borel set $\mathsf{C} \subset \{x, 0 \leq |x| \leq C\}$ satisfies the local Doeblin property (7), with $\lambda_{\mathsf{C}}(\cdot) = (2C)^{-1}\lambda^{\mathrm{Leb}}\left(\mathbb{1}_{\mathsf{C}}\cdot\right)$. Assumption (H1) is trivially satisfied. To check (H2), we set $\mathsf{K} = \mathsf{Y}$ and $V(x) = e^{c|x|}$ for some $c > 0$. The function $V^{-1}QV$ is locally bounded and $\lim_{|x|\to\infty} V^{-1}(x)QV(x) = 0$. Therefore, since $\sup_{\mathsf{X}\times\mathsf{Y}} g(x, y) \leq 1 \vee (2\pi\beta^2)^{-1/2}$, for any $\eta > 0$ one may choose a constant $C > 0$ large enough so that $\Upsilon_{\mathsf{C}^c}(y) \leq \eta\Upsilon_{\mathsf{X}}(y)$, where $\mathsf{C} \stackrel{\mathrm{def}}{=} \{0 \leq |x| \leq C\}$ and Υ_A is defined in (8). Therefore, (H2) is satisfied.

4.1.2 Application of Theorem 1

We now check conditions (12) to (14) of Theorem 1. Conditions (12) and (13) are obvious since $\mathsf{K} = \mathsf{Y}$ and $\sup_{\mathsf{Y}} \Upsilon_{\mathsf{X}} < \infty$. We now check (14) with $\mathsf{D} = \{0 \leq |x| \leq D\}$ and $\lambda_{\mathsf{D}}(\cdot) = (2D)^{-1}\lambda^{\mathrm{Leb}}\left(\mathbb{1}_{\mathsf{D}}\cdot\right)$ where the constant D is an arbitrary positive constant. $Q_\star(x, dy)$ is a

Gaussian density with mean $\phi_\star x$ and standard deviation $\sigma_\star$. Using standard arguments, $Q_\star$ is aperiodic, positive Harris with invariant distribution $\pi_\star$ which is a centered gaussian distribution with variance $\sigma_\star^2/(1-\phi_\star^2)$, and any compact set is petite. By the Jensen inequality, $\log \lambda_{\mathsf{D}}(g(\cdot,y)\mathbb{1}_{\mathsf{D}}) = \log \lambda_{\mathsf{D}}(g(\cdot,y)) \geq \lambda_{\mathsf{D}}(\log g(\cdot,y))$, which implies

$$\log \Psi_{\mathsf{D}}(y) = \log \lambda_{\mathsf{D}}(g(\cdot,y)) \geq \mathbb{1}\{y=0\} \log\left\{(2\pi\beta^2)^{-1/2}\int_D^\infty \mathrm{e}^{-v^2/2\beta^2}\lambda^{\mathrm{Leb}}(dv)\right\} \\ + \mathbb{1}\{y>0\}\left\{-(1/2)\log(2\pi\beta^2) - (12D\beta^2)^{-1}\left((D+y)^3+(D-y)^3\right)\right\} , \quad (33)$$

so that $\pi_\star \otimes G_\star([\log \Psi_{\mathsf{D}}]_-) < \infty$. Corollary 5 implies (14). Combining the results above, Theorem 1 therefore applies showing that (15) holds for any probability ν and ν' such that $\int \nu(dx)\mathrm{e}^{c|x|} + \int \nu'(dx)\mathrm{e}^{c|x|} < \infty$ for some $c > 0$.

4.1.3 Application of Theorem 3

We now consider the convergence of the expectation of the total variation distance at a polynomial rate. For all $p \geq 1$, there exists a constant C such that, for any $i \in \{0,1\}$, $\mathbb{E}_\star[Y_i^{2p}] \leq C(1+\mathbb{E}_\star[X_i^{2p}])$ which is finite since $\{X_i\}$ is Gaussian. Therefore,

$$\sup_{\mathsf{X}} (1+|x|^2)^{-p}\mathbb{E}_\star\left[\log g(x,Y_i)\right]_-^p < \infty , \quad (34)$$

which implies (24) since $V(x) = \exp(c|x|)$. By Proposition 7, there exists a constant C such that for any probability measure ν such that $\nu(V) < \infty$, $r_0(\nu,n) \leq Cn^{-p}\nu(V)$. Since $\sup_{\mathsf{Y}} \Upsilon_{\mathsf{X}} < \infty$, we may choose $M_1 > 0$ such that $M_1 > \sup_{\mathsf{Y}} \log \Upsilon_{\mathsf{X}}$; for this choice, $r_1(n) \equiv 0$, where $\{r_1(n)\}_{n\geq 0}$ is defined in (18). Since $\mathsf{K} = \mathsf{Y}$, $r_3(n) \equiv 0$, where $\{r_3(n)\}_{n\geq 0}$ is defined in (20). We now consider $\{r_2(n)\}_{n\geq 0}$ and apply Proposition 8. To that goal, we further assume that there exists $p_\star \geq 1$ such that $\nu_\star(|x|^{3p_\star+1}) < \infty$. It is easily seen that the drift condition (22) is satisfied with $V_\star(x) = 1+|x|^{3p_\star}$ and $f_\star \sim |x|^{3p_\star-1}$; furthermore, upon noting that $[\log \Psi_{\mathsf{D}}(y)]_- \sim |y|^2$, we have

$$\sup_{\mathsf{X}} f_\star^{-1} G_\star(\cdot, [\log \Psi_{\mathsf{D}}(y)]_-^{p_\star}) < \infty, \qquad \sup_{\mathsf{X}} f_\star^{-1} V_\star^{1-1/p_\star} G_\star(\cdot, [\log \Psi_{\mathsf{D}}(y)]_-) < \infty \ ,$$

thus proving $\limsup_{n\to\infty} n^{-(p_\star/2\vee 1)} r_2(n) = 0$. Therefore, by Theorem 3, the expectation $\mathbb{E}_\star\left(\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\right)$ goes to zero at the rate $n^{p_\star/2\vee 1}$ for any initial distributions ν, ν' such that $\int \{\nu(dx) + \nu'(dx)\} \exp(c|x|) < +\infty$.

The exponential decay can be proved similarly under the assumption that for some $c > 0$, $\int \nu_\star(dx) \exp(c|x|) < +\infty$; details are omitted.

4.2 Non-linear State-Space models

We consider the model (6) borrowed from (16). Assume that $\beta > 0$,

NLG(b, h) The functions b and h are locally bounded and

$$\lim_{|x|\to\infty} (|x + b(x)| - |x|) = -\infty \ . \tag{35}$$

NLG(σ) The noise variance is non-degenerated,

$$0 < \inf_{x\in\mathbb{R}^d} \inf_{\{\lambda\in\mathbb{R}^d, |\lambda|=1\}} \lambda^T \sigma(x)\sigma^T(x)\lambda \le \sup_{x\in\mathbb{R}^d} \sup_{\{\lambda\in\mathbb{R}^d, |\lambda|=1\}} \lambda^T \sigma(x)\sigma^T(x)\lambda < \infty \ . \tag{36}$$

The model is partially dominated with respect to the Lebesgue measure. The transition kernel $Q_{b,\sigma}$ and the likelihood $g_{h,\beta}$ are respectively given by

$$Q_{b,\sigma}(x, A) = (2\pi)^{-d/2} |\sigma(x)|^{-1} \int \exp\left(-(1/2)|x' - x - b(x)|^2_{\sigma(x)}\right) \mathbb{1}_A(x') \lambda^{\mathrm{Leb}}(dx') \ , \tag{37}$$

$$g_{h,\beta}(x, y) = (2\pi\beta^2)^{-\ell/2} \exp(-|y - h(x)|^2 / 2\beta^2) \ , \tag{38}$$

where $|u|^2_{\sigma(x)} = u^T [\sigma(x)\sigma^T(x)]^{-1} u$. As above, we set $Q = Q_{b,\sigma}$ and $g = g_{h,\beta}$.

Assume that $\{Y_k\}_{k\ge 0}$ are the observations of a non-linear Gaussian state space (6) with initial distribution $\nu_\star$ and 'parameters' $b_\star$, $h_\star$, $\sigma_\star$ and $\beta_\star$. We assume that $\beta_\star > 0$ and that the functions $b_\star$, $h_\star$ and $\sigma_\star$ satisfy NLG($b_\star, h_\star$)-NLG($\sigma_\star$), respectively, and

$$\limsup_{|x|\to\infty} |x|^{-1} \log |h_\star(x)| < \infty \; . \tag{39}$$

We denote by $Q_\star = Q_{b_\star,\sigma_\star}$, $G_\star = g_{h_\star,\beta_\star}\lambda^{\mathrm{Leb}}$ and $\mathbb{E}_\star = \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star}$ where $T_\star = Q_\star \otimes G_\star$.

4.2.1 Assumptions H1 and H2

Under NLG(b, h)-NLG(σ), every bounded Borel set in $\mathbb{R}^d$ is locally Doeblin in the sense given by (7). (H1) is trivial. Set $V(x) = \exp(c|x|)$, where c is a positive constant. The likelihood g is bounded by $(2\pi\beta^2)^{-\ell/2}$ and under NLG(σ), there exists a constant $M < \infty$ such that $V^{-1}(x)QV(x) \leq M \exp\left[c(|x + b(x)| - |x|)\right]$. Therefore, under NLG($b, h$)-NLG($\sigma$), for any $\eta > 0$, we may choose a constant C large enough such that $\Upsilon_{\mathsf{C}^c}(y) \leq \eta \Upsilon_{\mathsf{X}}(y)$ for any $y \in \mathsf{Y}$ where $\mathsf{C} = \{x \in \mathbb{R}^d, |x| \leq C\}$. Hence, assumption (H2) is satisfied with $\mathsf{K} = \mathsf{Y}$.

4.2.2 Application of Theorem 1

Condition (12) is trivial since $\mathsf{K} = \mathsf{Y}$. Condition (13) is obvious too since Υ_{X} is everywhere bounded. For (14), let us apply Corollary 5 and Proposition 6. $Q_\star$ is aperiodic, phi-irreducible and compact sets are petite. Set $\mathsf{D} = \{x \in \mathbb{R}^d, |x| \leq D\}$, where $D > 0$ and define $\lambda_{\mathsf{D}}(\cdot) = \lambda^{\mathrm{Leb}}(\mathbb{1}_{\mathsf{D}} \cdot)/\lambda^{\mathrm{Leb}}(\mathsf{D})$. Noting that $|y - h(x)|^2 \leq 2(|y|^2 + |h(x)|^2)$,

$$[\log g(x, y)]_- \leq \beta^{-2}|y|^2 + \beta^{-2}|h(x)|^2 + (\ell/2)\left[\log(2\pi\beta^2)\right]_+ \; . \tag{40}$$

Since the function h is locally bounded, $\sup_{\mathsf{D}} |h|^2 < \infty$ and (40) implies that

$$[\log \Psi_{\mathsf{D}}(y)]_- \leq \lambda_{\mathsf{D}}([\log g(\cdot, y)]_-) \leq \beta^{-2}|y|^2 + \beta^{-2} \sup_{\mathsf{D}} |h|^2 + (\ell/2)\left[\log(2\pi\beta^2)\right]_+ \; . \tag{41}$$

We set $V_\star(x) = \mathrm{e}^{c_\star |x|}$ we may find a compact (and thus petite) set $\mathsf{C}_\star$ and constants $\lambda_\star \in (0, 1)$ and $s_\star$ such that $Q_\star V_\star \leq \lambda_\star V_\star + s_\star \mathbb{1}_{\mathsf{C}_\star}$, so that (22) is satisfied with $f_\star = (1 - \lambda_\star)V_\star$.

Hence $Q_\star$ is positive Harris-recurrent and $\pi_\star(V_\star) < +\infty$. Furthermore, Eq. (41) implies that there exists a constant $C < \infty$ such that

$$G_\star\left(x, [\log \Psi_{\mathsf{D}}]_-\right) \leq C\left(1 + |h_\star(x)|^2\right) \leq C\left(1 + V_\star(x) \sup_{\mathsf{X}} V_\star^{-1} |h_\star|^2\right) . \tag{42}$$

The RHS is finite, provided $c_\star \geq 2 \limsup_{|x|\to\infty} |x|^{-1} \log |h_\star(x)|$ which we assume hereafter. Therefore, by Corollary 5 and Proposition 6, 1 applies: (15) holds for any initial probability measure such that $\int \mathrm{e}^{c|x|}\nu(dx) + \int \mathrm{e}^{c|x|}\nu'(dx) < +\infty$ for some $c > 0$.

4.2.3 Application of Theorem 3

We are willing to establish geometric rate of convergence and for that purpose we will use Proposition 7 and Proposition 10. We set $W(x) = c\{|x| - |x + b(x)|\} \vee 1$ and $W_\star(x) = c_\star\{|x| - |x + b_\star(x)|\} \vee 1$ and assume that

$$|h|^2 \in \mathcal{G}_W \quad \text{and} \quad |h_\star|^2 \in \mathcal{G}_{W_\star} . \tag{43}$$

$W_\star$ is unbounded and the level sets are petite for $Q_\star$. Furthermore, $V_\star(x) = \mathrm{e}^{c_\star |x|}$ where $c_\star > 0$ satisfies the multiplicative drift condition (27). Let $\lambda < \beta^2(2 \wedge \beta_\star^{-2})/4$. Since $\lambda\beta^{-2} < \beta_\star^{-2}/4$, Eq. (40) implies that there exists a constant $C < \infty$ such that for any integer i,

$$\mathbb{E}_\star\left[\mathrm{e}^{\lambda[\log g(x, Y_i)]_-}\right] \leq C \mathbb{E}_\star\left[\mathrm{e}^{2\lambda\beta^{-2}|h_\star(X_i)|^2}\right] \mathrm{e}^{\lambda\beta^{-2}|h(x)|^2} .$$

Since $\lambda \leq \beta^2/2$, Lemma 18 shows that $\sup_i \mathbb{E}_\star\left[\mathrm{e}^{2\lambda\beta^{-2}|h_\star(X_i)|^2}\right] < \infty$ provided $\nu_\star(V_\star) < +\infty$ which is henceforth assumed. Therefore, Proposition 7 applies, showing that there exists $\delta > 0$ such that for any probability measure ν such that $\nu(V) < \infty$, $r_0(\nu, n) \leq C\mathrm{e}^{-\delta n}\nu(V)$. As in Section 4.1, because Υ_{X} is bounded, we may choose M_1 large enough so that $r_1(n) \equiv 0$ (see (18)); similarly, since $\mathsf{K} = \mathsf{Y}$, $r_3(n) \equiv 0$. Eq. (41) implies that, for any $\lambda_\star$ small enough, $\log G_\star\left(\cdot, \mathrm{e}^{\lambda_\star[\log \Psi_{\mathsf{D}}]_-}\right) \in \mathcal{G}_{W_\star}$. Proposition 10 shows that $\limsup_{n\to\infty} n^{-1} \log r_2(n) < 0$. Hence Theorem 3 applies: for any initial distribution ν, ν'

such that $\int\{\nu(dx)+\nu'(dx)\}\exp(c|x|) < +\infty$, $\mathbb{E}_\star\left(\|\phi_{\nu,n}[Y_{0:n}] - \phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\right)$ goes to zero at a geometric rate.

4.3 Stochastic Volatility Model

As a final example, we consider the stochastic volatility (SV) model. In the canonical model in SV for discrete-time data (14; 15), the observations $\{Y_k\}_{k\geq 0}$ are the compounded returns and $\{X_k\}_{k\geq 0}$ is the log-volatility, which is assumed to follow a stationary auto-regression of order 1, *i.e.*

$$\begin{cases} X_{k+1} = \phi X_k + \sigma\zeta_k \ , \\ Y_k = \beta\exp(X_k/2)\varepsilon_k \ , \end{cases} \tag{44}$$

where $\{(\zeta_k, \varepsilon_k)\}_{k\geq 0}$ is a i.i.d. sequence of standard Gaussian vectors, $|\phi| < 1$, $\sigma > 0$ and $\beta > 0$. Here $\mathsf{X} = \mathsf{Y} = \mathbb{R}$ and $\mathcal{X}$ and $\mathcal{Y}$ are the Borel sigma-fields. The model is partially dominated with respect to the Lebesgue measure. The transition kernel $Q_{\phi,\sigma}$ and the likelihood g_β are respectively given by

$$Q_{\phi,\sigma}(x, A) = (2\pi\sigma^2)^{-1/2}\int \exp(-1/(2\sigma^2)(x' - \phi x)^2\mathbb{1}_A(x')\lambda^{\mathrm{Leb}}(dx') \ , \tag{45}$$

$$g_\beta(x, y) = (2\pi\beta^2)^{-1/2}\exp\left(-y^2\exp(-x)/2\beta^2 - x/2\right) \ . \tag{46}$$

We denote $Q = Q_{\phi,\sigma}$ and $g = g_\beta$.

We assume that $\{Y_k\}_{k\geq 0}$ are the observations of the stochastic volatility model (44) with initial distribution $\nu_\star$ and parameters $|\phi_\star| < 1$, $\sigma_\star > 0$, and $\beta_\star > 0$. We denote as above $Q_\star = Q_{\phi_\star,\sigma_\star}$, $G_\star = g_{\beta_\star}\lambda^{\mathrm{Leb}}$, $T_\star = Q_\star \otimes G_\star$ and $\mathbb{E}_\star = \mathbb{E}^{T_\star}_{\nu_\star\otimes G_\star}$.

4.3.1 Assumptions H1 and H2

As in example 4.1, every bounded Borel set is locally Doeblin in the sense of (7). Assumption (H1) is satisfied but the likelihood is not uniformly bounded over $\mathsf{X} \times \mathsf{Y}$; nevertheless it is easily seen that $\sup_{x \in \mathsf{X}} g(x,y) \le (2\pi \mathrm{e})^{-1/2} |y|^{-1}$. We set $\mathsf{K} = \mathbb{R}$ and put $V(x) = \mathrm{e}^{c|x|}$ where c is positive; as in Example 4.1, $QV(\cdot)/V(\cdot)$ is locally bounded and $\lim_{|x| \to \infty} QV(x)/V(x) = 0$, showing that assumption (H2) is fulfilled.

4.3.2 Application of Theorem 1

The Markov kernel $Q_\star$ is positive recurrent, geometrically ergodic and its stationary distribution $\pi_\star$ is Gaussian with mean 0 and variance $\sigma_\star^2/(1-\phi_\star^2)$. Note that there exists a constant $C < \infty$ such that for all $y \in \mathsf{Y}$, $[\log \Upsilon_{\mathsf{X}}(y)]_+ \le C \,|\log |y||$, which implies that $G_\star(x, [\log \Upsilon_{\mathsf{X}}]_+) < C + |x|/2$ for some constant $C < \infty$. This implies that $\pi_\star \otimes G_\star([\log \Upsilon_{\mathsf{X}}]_+) < \infty$ and Corollary 5 implies (13). Set $\mathsf{D} = \{x, |x| \le D\}$ where $D > 0$ and let $\lambda_{\mathsf{D}}(\cdot) = \lambda^{\mathrm{Leb}}(\mathbb{1}_{\mathsf{D}} \cdot)/\lambda^{\mathrm{Leb}}(\mathsf{D})$. By the Jensen inequality,

$$\log \Psi_{\mathsf{D}}(y) \ge \lambda_{\mathsf{D}}(\log g(\cdot, y)) = -(1/2)\log(2\pi\beta^2) - y^2 \, \mathrm{sh}(D)/[2\beta^2 D] \;,$$

showing that there exists a constant $C < \infty$ such that $[\log \Psi_{\mathsf{D}}(y)]_- \le C(1+y^2)$. Therefore, $G_\star(x, [\log \Psi_{\mathsf{D}}]_-) \le C(1 + \beta^2 \mathrm{e}^x)$. The conditions of Corollary 5 are satisfied, showing that (14) holds. As a result, (15) holds for any initial distributions ν and ν' such that $\int \nu(dx) \exp(c|x|) + \int \nu'(dx) \exp(c|x|) < \infty$.

The problem of computing the convergence rates can be addressed as in the other examples.

5 Proof of Theorems 1 and 3

Before proving the main results, some additional definitions are needed. A function $\bar{f}$ defined on $\bar{\mathsf{X}} \overset{\mathrm{def}}{=} \mathsf{X} \times \mathsf{X}$ is said to be *symmetric* if for all $(x,x') \in \bar{\mathsf{X}}$, $f(x,x') = f(x',x)$. An unnormalised transition kernel P on $(\bar{\mathsf{X}}, \bar{\mathcal{X}})$, where $\bar{\mathcal{X}} = \mathcal{X} \otimes \mathcal{X}$ is said to be *symmetric* if for all (x, x') in $\bar{\mathsf{X}}$ and any positive symmetric function f, $P[(x, x'), f] = P[(x', x), f]$. For P a Markov kernel on $(\mathsf{X}, \mathcal{X})$, we denote by $\bar{P}$ the transition kernel on $(\bar{\mathsf{X}}, \bar{\mathcal{X}})$ defined, for any $(x, x') \in \bar{\mathsf{X}}$ and A, $A' \in \mathcal{X}$, by

$$\bar{P}[(x, x'), A \times A'] = P(x, A) P(x', A') \; . \tag{47}$$

For any $A \in \mathcal{X}$, and ν and ν' two probability distributions on $(\mathsf{X}, \mathcal{X})$ the difference $\phi_{\nu,n}[y_{0:n}](A) - \phi_{\nu',n}[y_{0:n}](A)$ may be expressed as

$$\begin{aligned} &\phi_{\nu,n}[y_{0:n}](A) - \phi_{\nu',n}[y_{0:n}](A) \\ &= \frac{\mathbb{E}^Q_\nu\left[\prod_{i=0}^n g(X_i, y_i) \mathbb{1}_A(X_n)\right]}{\mathbb{E}^Q_\nu\left[\prod_{i=0}^n g(X_i, y_i)\right]} - \frac{\mathbb{E}^Q_{\nu'}\left[\prod_{i=0}^n g(X_i, y_i) \mathbb{1}_A(X_n)\right]}{\mathbb{E}^Q_{\nu'}\left[\prod_{i=0}^n g(X_i, y_i)\right]} \\ &= \frac{\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \bar{g}(X_i, X'_i, y_i) \mathbb{1}_A(X_n)\right] - \mathbb{E}^{\bar{Q}}_{\nu'\otimes\nu}\left[\prod_{i=0}^n \bar{g}(X_i, X'_i, y_i) \mathbb{1}_A(X_n)\right]}{\mathbb{E}^Q_\nu\left[\prod_{i=0}^n g(X_i, y_i)\right] \mathbb{E}^Q_{\nu'}\left[\prod_{i=0}^n g(X_i, y_i)\right]} \; , \end{aligned} \tag{48}$$

where $\bar{g}(x, x', y) = g(x, y) g(x', y)$. The idea of writing the difference using a pair of independent processes has been apparently introduced in (3); this approach is central in the work of (16). We consider separately the numerator and the denominator of Eq. (48). For the numerator, the path of the independent processes is decomposed along the successive visits to $\mathsf{C} \times \mathsf{C}$ as done in (16).

Proposition 12 *Let C be a LD-set and ν and ν' be two probability distributions on $(\mathsf{X}, \mathcal{X})$. For any integer n and functions $g_i \in \mathbb{B}_+(\mathsf{X})$, $i = 0, \ldots, n$, such that $\mathbb{E}^Q_\nu\left[\prod_{i=0}^n g_i(X_i)\right] < \infty$ and $\mathbb{E}^Q_{\nu'}\left[\prod_{i=0}^n g_i(X_i)\right] < \infty$, define*

$$\begin{aligned} &\Delta_n(\nu, \nu', \{g_i\}_{i=0}^n) \\ &= \sup_{A \in \mathcal{X}} \left| \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \bar{g}_i(X_i, X'_i) 1_A(X_n)\right] - \mathbb{E}^{\bar{Q}}_{\nu'\otimes\nu}\left[\prod_{i=0}^n \bar{g}_i(X_i, X'_i) 1_A(X_n)\right] \right| \; , \end{aligned} \tag{49}$$

where $\bar{g}_i(x, x') = g_i(x) g_i(x')$. Then,

$$\Delta_n(\nu, \nu', \{g_i\}_{i=1}^n) \le \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \bar{g}_i(X_i, X'_i) \rho_{\mathsf{C}}^{N_{\mathsf{C},n}}\right] \; , \tag{50}$$

where $\bar{Q}$ is defined as in (47) *and*

$$N_{\mathsf{C},n} \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} \mathbb{1}_{\mathsf{C}\times\mathsf{C}}(X_i, X'_i)\mathbb{1}_{\mathsf{C}\times\mathsf{C}}(X_{i+1}, X'_{i+1}) \;, \tag{51}$$

$$\rho_{\mathsf{C}} \stackrel{\text{def}}{=} 1 - \left(\epsilon^-_{\mathsf{C}}/\epsilon^+_{\mathsf{C}}\right)^2 \;. \tag{52}$$

PROOF. Put $\bar{x} = (x, x')$, $\bar{g}_i(\bar{x}) = g_i(x)g_i(x')$, $\bar{\mathsf{C}} \stackrel{\text{def}}{=} \mathsf{C} \times \mathsf{C}$, and $\bar{\lambda}_{\bar{\mathsf{C}}} \stackrel{\text{def}}{=} \lambda_{\mathsf{C}} \otimes \lambda_{\mathsf{C}}$. We stress that the kernels that will be defined along this proof may be unnormalized. Since C is a locally Doeblin set, we have for any measurable positive function $\bar{f}$ on $(\bar{\mathsf{X}}, \bar{\mathcal{X}})$,

$$(\epsilon^-_{\mathsf{C}})^2\bar{\lambda}_{\bar{\mathsf{C}}}(\mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) \le \bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) \le (\epsilon^+_{\mathsf{C}})^2\bar{\lambda}_{\bar{\mathsf{C}}}(\mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) \;, \quad \text{for all } \bar{x} \in \bar{\mathsf{C}} \;. \tag{53}$$

Define the unnormalised kernel $\bar{Q}_0$ and $\bar{Q}_1$ by

$$\bar{Q}_0(\bar{x}, \bar{f}) \stackrel{\text{def}}{=} \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})(\epsilon^-_{\mathsf{C}})^2\bar{\lambda}_{\bar{\mathsf{C}}}(\mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) \tag{54}$$

$$\bar{Q}_1(\bar{x}, \bar{f}) \stackrel{\text{def}}{=} \bar{Q}(\bar{x}, \bar{f}) - \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})(\epsilon^-_{\mathsf{C}})^2\bar{\lambda}_{\bar{\mathsf{C}}}(\mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) = \bar{Q}(\bar{x}, \bar{f}) - \bar{Q}_0(\bar{x}, \bar{f}) \;. \tag{55}$$

Eq. (53) implies that, for all $\bar{x} \in \bar{\mathsf{C}}$, $0 \le \bar{Q}_1(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) \le \rho_{\mathsf{C}}\bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}}\bar{f})$. It then follows using straightforward algebra that,

$$\begin{aligned} \bar{Q}_1(\bar{x}, \bar{f}) &= \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})\bar{Q}_1(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})\bar{Q}_1(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}^c}\bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x})\bar{Q}_1(\bar{x}, \bar{f}) \\ &\le \rho_{\mathsf{C}}\mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})\bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}}\bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})\bar{Q}(\bar{x}, \mathbb{1}_{\bar{\mathsf{C}}^c}\bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x})\bar{Q}(\bar{x}, \bar{f}) \\ &\le \bar{Q}(\bar{x}, \rho_{\mathsf{C}}^{\mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})\mathbb{1}_{\bar{\mathsf{C}}}}\bar{f}) \;. \end{aligned} \tag{56}$$

We write $\Delta_n(\nu, \nu', \{g_i\}_{i=0}^n) = \sup_{A\in\mathcal{X}} |\Delta_n(A)|$ where

$$\Delta_n(A) \stackrel{\text{def}}{=} \nu \otimes \nu' \left(\bar{g}_0\bar{Q}\bar{g}_1 \cdots \bar{g}_{n-1}\bar{Q}\bar{g}_n\mathbb{1}_{A\times\mathsf{X}}\right) - \nu' \otimes \nu \left(\bar{g}_0\bar{Q}\bar{g}_1 \cdots \bar{g}_{n-1}\bar{Q}\bar{g}_n\mathbb{1}_{A\times\mathsf{X}}\right) \;. \tag{57}$$

Note that $\Delta_n(A)$ may be decomposed as $\Delta_n(A) = \sum_{t_{0:n-1} \in \{0,1\}^n} \Delta(A, t_{0:n-1})$ where

$$\Delta_n(A, t_{0:n-1}) \stackrel{\text{def}}{=} \nu \otimes \nu' \left(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \cdots \bar{g}_{n-1} \bar{Q}_{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}} \right) \\ - \nu' \otimes \nu \left(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \cdots \bar{g}_{n-1} \bar{Q}_{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times \mathsf{X}} \right) \; .$$

Note that, for any $t_{0:n-1} \in \{0,1\}^n$ and any sets $A, B \in \mathcal{X}$,

$$\nu' \otimes \nu \left(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \cdots \bar{g}_{n-1} \bar{Q}_{t_{n-1}} \bar{g}_n \mathbb{1}_{A \times B} \right) \\ = \nu \otimes \nu' \left(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \cdots \bar{g}_{n-1} \bar{Q}_{t_{n-1}} \bar{g}_n \mathbb{1}_{B \times A} \right) \; . \quad (58)$$

First assume that there exists an index $i \geq 0$ such that $t_i = 0$ then,

$$\begin{aligned} &\nu \otimes \nu' \left(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \cdots \bar{g}_{n-1} \bar{Q}_{t_{n-1}} \bar{g}_n 1_{A \times \mathsf{X}} \right) \\ &\quad = \nu \otimes \nu' \left(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \cdots \bar{Q}_{t_{i-1}} \bar{g}_i \mathbb{1}_{\bar{\mathsf{C}}} \right) \times (\epsilon_{\mathsf{C}}^-)^2 \bar{\lambda}_{\mathsf{C}} \left(\mathbb{1}_{\bar{\mathsf{C}}} \bar{g}_{i+1} \bar{Q}_{t_{i+1}} \ldots \bar{g}_{n-1} \bar{Q}_{t_{n-1}} \bar{g}_n 1_{A \times \mathsf{X}} \right) \\ &\quad = \nu' \otimes \nu \left(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \cdots \bar{Q}_{t_{i-1}} \bar{g}_i \mathbb{1}_{\bar{\mathsf{C}}} \right) \times (\epsilon_{\mathsf{C}}^-)^2 \bar{\lambda}_{\mathsf{C}} \left(\mathbb{1}_{\bar{\mathsf{C}}} \bar{g}_{i+1} \bar{Q}_{t_{i+1}} \ldots \bar{g}_{n-1} \bar{Q}_{t_{n-1}} \bar{g}_n 1_{A \times \mathsf{X}} \right) \end{aligned}$$

by (58). Thus, $\Delta_n(A, t_{0:n-1})$ is equal to 0 except if for all i, $t_i = 1$, and (58) finally implies

$$\Delta_n(A) = \nu \otimes \nu' \left[\bar{g}_0 \bar{Q}_1 \bar{g}_1 \cdots \bar{g}_{n-1} \bar{Q}_1 \bar{g}_n (\mathbb{1}_{A \times \mathsf{X}} - \mathbb{1}_{\mathsf{X} \times A}) \right] \; .$$

Using (56), we have

$$\Delta_n(\nu, \nu', \{g_i\}_{i=0}^n) \leq \nu \otimes \nu' \left(\bar{g}_0 \bar{Q}_1 \bar{g}_1 \cdots \bar{g}_{n-1} \bar{Q}_1 \bar{g}_n \right) \\ \leq \mathbb{E}_{\nu \otimes \nu'}^{\bar{Q}} \left[\prod_{i=0}^n \bar{g}_i(\bar{X}_i) \rho_{\mathsf{C}}^{\sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{C}}}(\bar{X}_i) \mathbb{1}_{\bar{\mathsf{C}}}(\bar{X}_{i+1})} \right] ,$$

where the last equality is straightforward to establish by induction on n. The proof is completed.

Remark 13 *If the whole state space* X *is a locally Doeblin, then one may take* $\mathsf{C} = \mathsf{X}$ *in the previous expression. Since* $N_{\mathsf{X},n} = n$, (48) *and the previous proposition therefore imply the uniform ergodicity of the filtering distribution, for any initial distribution* ν *and* ν', *and any sequence* $y_{0:n}$*,* $\|\phi_{\nu,n}[y_{0:n}] - \phi_{\nu',n}[y_{0:n}]\|_{\mathrm{TV}} \leq \rho_{\mathsf{X}}^n$ *where* $\rho_{\mathsf{X}} \stackrel{\mathrm{def}}{=} 1 - (\epsilon_{\mathsf{X}}^- / \epsilon_{\mathsf{X}}^+)^2$*; see (2) and (10).*

We consider now the denominator of (48). A lower bound for the denominator has been computed in (4, Lemma 2.2). This is obtained by using a change of measure ideas. We use here a more straightforward argument.

Proposition 14 *For any LD-set* $\mathsf{C} \in \mathcal{X}$*,* $n \geq 1$ *and any functions* $g_i \in \mathbb{B}_+(\mathsf{X})$*,* $i = 0, \ldots, n$*,*

$$\mathbb{E}_\nu^Q \left[\prod_{i=0}^n g_i(X_i) \right] \geq (\epsilon_{\mathsf{C}}^-)^{n-1} \nu(g_0 Q g_1 \mathbb{1}_{\mathsf{C}}) \prod_{i=2}^n \lambda_{\mathsf{C}}(g_i \mathbb{1}_{\mathsf{C}}) \; .$$

PROOF. The proof follows immediately from

$$\mathbb{E}_\nu^Q \left[\prod_{i=0}^n g_i(X_i) \right] \geq \mathbb{E}_\nu^Q \left[g_0(X_0) \prod_{i=1}^n g_i(X_i) \mathbb{1}_{\mathsf{C}}(X_i) \right] \; ,$$

and the minorization condition (7).

By combining Propositions 12 and 14, we can obtain an explicit bound for the total variation distance $\|\phi_{\nu,n}[y_{0:n}] - \phi_{\nu',n}[y_{0:n}]\|_{\mathrm{TV}}$.

Lemma 15 *Let* $\beta \in (0, 1)$*. Then, for any LD-sets* $\mathsf{C} \subseteq \mathsf{X}$ *and* $\mathsf{D} \subseteq \mathsf{X}$*, any initial probability measures* ν *and* ν'*, any function* $V : \mathsf{X} \to [1, \infty)$*,*

$$\begin{aligned} \|\phi_{\nu,n}[y_{0:n}] - \phi_{\nu',n}[y_{0:n}]\|_{\mathrm{TV}} &\leq \rho_{\mathsf{C}}^{\beta n} \\ &+ \frac{\prod_{i=0}^n \Upsilon_{\mathsf{X}}(y_i) \ \max_{\mathcal{I} \subset \{0,\ldots,n\}, |\mathcal{I}| = a_n} \prod_{i \in \mathcal{I}} \Upsilon_{\mathsf{C}^c}(y_i) \prod_{i \notin \mathcal{I}} \Upsilon_{\mathsf{X}}(y_i)}{(\epsilon_{\mathsf{D}}^-)^{2(n-1)} \Phi_{\nu,\mathsf{D}}(y_0, y_1) \ \Phi_{\nu',\mathsf{D}}(y_0, y_1) \prod_{i=2}^n \Psi_{\mathsf{D}}^2(y_i)} \nu(V) \nu'(V) \; , \end{aligned}$$

where $a_n \stackrel{\mathrm{def}}{=} \lfloor n(1-\beta)/2 \rfloor$*,* $|\mathcal{I}|$ *is the cardinal of the set* $\mathcal{I}$ *and the functions* $\Phi_{\nu,\mathsf{D}}$ *and* Ψ_{D} *are defined in* (10) *and* (11)*, respectively.*

PROOF. Eq. (50) implies that for any $\beta \in (0,1)$,

$$\Delta_n(\nu, \tilde{\nu}, \{g_i\}_{i=0}^n) \leq \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \bar{g}(\bar{X}_i, y_i)\rho_{\mathsf{C}}^{N_{\mathsf{C},n}} \mathbb{1}\{N_{\mathsf{C},n} \geq \beta n\}\right] \\ + \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \bar{g}(\bar{X}_i, y_i)\rho_{\mathsf{C}}^{N_{\mathsf{C},n}} \mathbb{1}\{N_{\mathsf{C},n} < \beta n\}\right] .$$

The first term in the RHS is bounded by $\rho_{\mathsf{C}}^{\beta n}\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \bar{g}(\bar{X}_i, y_i)\right]$. We now consider the second term. For any set $\mathsf{A} \in \bar{\mathcal{X}}$, denote by $M_{\mathsf{A},n}$ the number of visits of $\{\bar{X}_k\}_{k\geq 0}$ to the set A before n. By Lemma 17, the condition $N_{\mathsf{C},n} < \beta n$ implies that $M_{\bar{\mathsf{C}},n} < n(1+\beta)/2$ and $M_{\bar{\mathsf{C}}^c,n} \geq a_n$. Note that for any $\bar{x} \in \bar{\mathsf{X}}$ and $y \in \mathsf{Y}$,

$$\bar{g}(\bar{x}, y)\bar{Q}\bar{V}(\bar{x}) \leq [A(y)]^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x})}[B(y)]^{\mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})}\bar{V}(\bar{x}) , \tag{59}$$

where we have set $\bar{V}(\bar{x}) \stackrel{\text{def}}{=} V(x)V(x')$, $A(y) \stackrel{\text{def}}{=} \sup_{\bar{x}\in\bar{\mathsf{C}}^c} \bar{g}(\bar{x}, y)\bar{V}^{-1}(\bar{x})\bar{Q}\bar{V}(\bar{x})$, and $B(y) \stackrel{\text{def}}{=} \sup_{\bar{x}\in\bar{\mathsf{X}}} \bar{g}(\bar{x}, y)\bar{V}^{-1}(\bar{x})\bar{Q}\bar{V}(\bar{x})$. Consider the process

$$V_0 = \bar{V}(\bar{X}_0), \quad \text{and} \quad V_n \stackrel{\text{def}}{=} \left\{\prod_{i=0}^{n-1} \frac{\bar{g}(\bar{X}_i, y_i)}{[A(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{X}_i)}[B(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}}(\bar{X}_i)}}\right\} \bar{V}(\bar{X}_n) \, , n \geq 1 \, , \tag{60}$$

where by convention we have set $0/0 = 0$ (to deal with cases where either $A(y) = 0$ or $B(y) = 0$). The process $\{V_n\}_{n\geq 0}$ is a $\mathcal{F}$-super-martingale, where $\mathcal{F} = \{\mathcal{F}_n\}$ is the natural filtration of the process $\{\bar{X}_k\}_{k\geq 0}$, $\mathcal{F}_n \stackrel{\text{def}}{=} \sigma(\bar{X}_0, \ldots, \bar{X}_n)$. Denote by τ_{a_n} the a_n-th return time to the set $\bar{\mathsf{C}}^c$. On the event $\{M_{\bar{\mathsf{C}}^c,n} \geq a_n\}$, $\tau_{a_n} \leq n$, using that $A(y) \leq B(y)$

$$\prod_{i=0}^n [A(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{X}_i)}[B(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}}(\bar{X}_i)} \leq \\ \prod_{i=0}^{\tau_n} [A(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{X}_i)}[B(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}}(\bar{X}_i)} \prod_{i=\tau_{a_n}+1}^n B(y_i) \leq C(y_{0:n})$$

where $C(y_{0:n}) \stackrel{\text{def}}{=} \max_{\mathcal{I}\subset\{0,\ldots,n\}, |\mathcal{I}|=a_n} \prod_{i\in\mathcal{I}} A(y_i) \prod_{i\notin\mathcal{I}} B(y_i)$. Therefore,

$$\begin{aligned}
\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}&\left[\prod_{i=0}^n \bar{g}(\bar{X}_i, y_i)\mathbb{1}\{N_{\mathsf{C},n} < \beta n\}\right] \leq \mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \bar{g}(\bar{X}_i, y_i)\mathbb{1}\{M_{\bar{\mathsf{C}}^c,n} \geq a_n\}\right] \\
&\leq C(y_{0:n})\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^n \frac{\bar{g}(\bar{X}_i, y_i)}{[A(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{X}_i)}[B(y_i)]^{\mathbb{1}_{\bar{\mathsf{C}}}(\bar{X}_i)}}\bar{V}(\bar{X}_{n+1})\right] \\
&= C(y_{0:n})\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}[V_{n+1}] \, .
\end{aligned}$$

The super-martingale inequality therefore implies

$$\mathbb{E}^{\bar{Q}}_{\nu\otimes\nu'}\left[\prod_{i=0}^{n}\bar{g}(\bar{X}_i,y_i)\mathbb{1}\{N_{\mathsf{C},n}<\beta n\}\right]\leq C(y_{0:n})\nu(V)\nu'(V)\;,$$

and the proof follows from (48) and Proposition 14, using that $A(y)\leq\Upsilon_{\mathsf{X}}(y)\Upsilon_{\mathsf{C}^c}(y)$ and $B(y)=\Upsilon^2_{\mathsf{X}}(y)$, where $\Upsilon_A(y)$ is defined in (8).

Corollary 16 *Assume (H2). Let* D *be a LD-set, and* γ *and* β *be constants satisfying* $\gamma\in(0,1)$ *and* $\beta\in(0,\gamma)$. *Then, for any* $\eta\in(0,1)$ *there exists a LD-set* C *such that, for any sequence* $y_{0:n}\in\mathsf{Y}^{n+1}$ *satisfying* $\sum_{i=0}^{n}\mathbb{1}_{\mathsf{K}}(y_i)\geq(1+\gamma)n/2$, *any initial probability measures* ν *and* ν', *and any* $n\geq 1$,

$$\|\phi_{\nu,n}[y_{0:n}]-\phi_{\nu',n}[y_{0:n}]\|_{\mathrm{TV}}\leq\rho_{\mathsf{C}}^{\beta n}+\frac{\eta^{(\gamma-\beta)n/2}\prod_{i=0}^{n}\Upsilon^2_{\mathsf{X}}(y_i)}{(\epsilon^-_{\mathsf{D}})^{2(n-1)}\Phi_{\nu,\mathsf{D}}(y_0,y_1)\;\Phi_{\nu',\mathsf{D}}(y_0,y_1)\prod_{i=2}^{n}\Psi^2_{\mathsf{D}}(y_i)}\,\nu(V)\nu'(V)\;,$$

where ρ_{C} , $\Phi_{\nu,\mathsf{D}}$ *and* Ψ_{D} *are defined in* (52), (10) *and* (11), *respectively.*

PROOF. [Proof of Theorem 1] The conditions (13) and (14) imply that

$$\limsup_{n\to\infty}\;\exp(-2Mn)\prod_{i=0}^{n}\Upsilon^2_{\mathsf{X}}(Y_i)\leq 1\quad\text{and}\quad\limsup_{n\to\infty}\;\exp(-2Mn)\prod_{i=0}^{n}\Psi^{-2}_{\mathsf{D}}(Y_i)\leq 1\;.$$

Condition (H1) and $\nu Q\mathbb{1}_{\mathsf{D}}>0$ implies that $\Phi_{\nu,\mathsf{D}}(y_0,y_1)>0$ for any $(y_0,y_1)\in\mathsf{Y}^2$. We then choose η small enough so that

$$\lim_{n\to\infty}\eta^{(\gamma-\beta)n/2}\exp(4Mn)(\epsilon^-_{\mathsf{D}})^{-2(n-1)}=0\;.$$

The proof follows from Corollary 16.

PROOF. [Proof of Theorem 3] Note that for any $\alpha\in(0,1)$ and any integer n,

$$\mathbb{E}_\star[\|\phi_{\nu,n}[Y_{0:n}]-\phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}]\leq\alpha^n+\mathbb{P}_\star[\|\phi_{\nu,n}[Y_{0:n}]-\phi_{\nu',n}[Y_{0:n}]\|_{\mathrm{TV}}\geq\alpha^n]\;.$$

Consider now the second term in the RHS of the previous equation. Denote Ω_n the event

$$\Omega_n \stackrel{\text{def}}{=} \Big\{\log \Phi_{\nu,\mathsf{D}}(Y_0,Y_1) > -M_0 n\ , \log \Phi_{\nu',\mathsf{D}}(Y_0,Y_1) > -M_0 n\ ,\\ \sum_{i=0}^{n} \log \Upsilon_{\mathsf{X}}(Y_i) < M_1 n, \sum_{i=2}^{n} \log \Psi_{\mathsf{D}}(Y_i) > -M_2 n, \sum_{i=1}^{n} \mathbb{1}_{\mathsf{K}}(Y_i) > n(1+\gamma)/2\Big\}\ .$$

Clearly, $\mathbb{P}_\star(\Omega_n^c) \leq \sum_{i=1}^3 r_i(n) + r_0(\nu,n) + r_0(\nu',n)$ where $\{r_i(n)\}_{n\geq 0}$ and $\{r_0(\nu,n)\}_{n\geq 0}$ are defined in Eqs. (17)-(20). On the event Ω_n,

$$\Phi_{\nu,\mathsf{D}}^{-1}(Y_0,Y_1)\ \Phi_{\nu',\mathsf{D}}^{-1}(Y_0,Y_1)\ \prod_{i=2}^{n} \Psi_{\mathsf{D}}^{-2}(Y_i)\ \prod_{i=0}^{n} \Upsilon_{\mathsf{X}}(Y_i) \leq \mathrm{e}^{2n\sum_{i=0}^2 M_i}\ .$$

One may choose $\eta > 0$ small enough and $\varrho \in (0,1)$ so that, for any n,

$$\eta^{(\gamma-\beta)n/2} \mathrm{e}^{2n\sum_{i=0}^2 M_i} (\epsilon_{\mathsf{D}}^-)^{-2(n-1)} \leq \varrho^n\ .$$

The proof then follows from Corollary 16.

6 Proof of Propositions 7, 8, and 10

PROOF. [Proof of Proposition 7] By the Jensen inequality with the function $u \mapsto [\log(u)]_-^p$, we obtain that for any $p \geq 1$,

$$[\log \Phi_{\nu,\mathsf{D}}(Y_0,Y_1) - \log(\nu Q \mathbb{1}_{\mathsf{D}})]_-^p \\ \leq 2^{p-1} (\nu Q \mathbb{1}_{\mathsf{D}})^{-1} \iint \nu(dx_0) Q(x_0,dx_1) \mathbb{1}_{\mathsf{D}}(x_1)\ \sum_{i=0}^{1} [\log g(x_i,Y_i)]_-^p\ , \quad (61)$$

which implies by the Fubini theorem,

$$\mathbb{E}_\star \left\{[\log \Phi_{\nu,\mathsf{D}}(Y_0,Y_1) - \log(\nu Q \mathbb{1}_{\mathsf{D}})]_-^p\right\} \\ \leq 2^{p-1} (\nu Q \mathbb{1}_{\mathsf{D}})^{-1} \iint \nu(dx_0) Q(x_0,dx_1) \sum_{i=0}^{1} \mathbb{E}_\star [\log g(x_i,Y_i)]_-^p\ .$$

Since $\sup_{\mathsf{X}} V^{-1}\mathbb{E}_\star[\log g(\cdot, Y_i)]_-^p < \infty$, and $\sup_{\mathsf{X}} V^{-1}QV < \infty$,

$$\iint \nu(dx_0)Q(x_0, dx_1) \sum_{i=0}^{1} \mathbb{E}_\star[\log g(x_i, Y_i)]_-^p \\ \leq \nu(V) \left\{ \sup_{i=0,1} \sup_{\mathsf{X}} V^{-1}\mathbb{E}_\star[\log g(\cdot, Y_i)]_-^p \right\} (1 + \sup_{\mathsf{X}} V^{-1}QV) \ . \quad (62)$$

Similarly, for $\lambda > 0$, using the Jensen inequality with $u \mapsto \exp[(\lambda/2)[\log u]_-]$ and the Fubini Theorem, we have

$$\mathbb{E}_\star \left[\exp\left((\lambda/2)[\log \Phi_{\nu,\mathsf{D}}(Y_0, Y_1) - \log(\nu Q \mathbb{1}_{\mathsf{D}})]_-\right)\right] \leq (\nu Q \mathbb{1}_{\mathsf{D}})^{-1} \\ \times \iint \nu(dx_0)Q(x_0, dx_1) \mathbb{E}_\star^{1/2}\left[\exp(\lambda[\log g(x_0, Y_0)]_-)\right] \mathbb{E}_\star^{1/2}\left[\exp(\lambda[\log g(x_1, Y_1)]_-)\right] \ ,$$

and the proof follows since $\sup_{\mathsf{X}} V^{-1/2}QV^{1/2} < \infty$.

PROOF. [Proof of Proposition 8] Let φ be a non negative function on Y. Assume that $\sup_{\mathsf{X}} U_\star^{-1} G_\star(\cdot, \varphi^p) < \infty$. Proposition 6 shows that $\pi_\star[G_\star(\cdot, \varphi^p)] < \infty$. Without loss of generality, we assume that $\pi_\star[G_\star(\cdot, \varphi)] = 0$. For any $p \geq 1$,

$$\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} \varphi(Y_i) \right|^p \\ \leq 2^{p-1} \left(\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} \{\varphi(Y_i) - G_\star(X_i, \varphi)\} \right|^p + \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} G_\star(X_i, \varphi) \right|^p \right) \ . \quad (63)$$

Since conditionally to $X_{0:n}$ the random variables $Y_{0:n}$ are independent, we may apply the Marcinkiewicz-Zygmund inequality (13, Inequality 2.6.18 p. 82), showing that there exists a constant $c(p)$ depending only on p such that

$$\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} \{\varphi(Y_i) - G_\star(X_i, \varphi)\} \right|^p \leq c(p) \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left(\sum_{i=0}^{n} |\varphi(Y_i) - G_\star(X_i, \varphi)|^2 \right)^{p/2} \ .$$

If $1 \leq p \leq 2$,

$$\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} \{\varphi(Y_i) - G_\star(X_i, \varphi)\} \right|^p \leq c(p) \sum_{i=0}^{n} \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left[|\varphi(Y_i) - G_\star(X_i, \varphi)|^p \right] \\ \leq 2^p c(p) \sum_{i=0}^{n} \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left[G_\star(X_i, |\varphi|^p) \right] \ .$$

If $p \geq 2$, the Minkowski inequality yields

$$\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} \{\varphi(Y_i) - G_\star(X_i, \varphi)\} \right|^p \leq c(p) \left(\sum_{i=0}^{n} \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left[|\varphi(Y_i) - G_\star(X_i, \varphi)|^p \right]^{2/p} \right)^{p/2}$$
$$\leq 2^p c(p) n^{p/2-1} \sum_{i=0}^{n} \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left[G_\star(X_i, |\varphi|^p) \right] \; .$$

The f-norm ergodic theorem (20, Theorem 14.0.1) implies that there exists a constant $C < \infty$, such that for any initial probability measure $\nu_\star$,

$$\sum_{i=0}^{n} \mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left[G_\star(X_i, |\varphi|^p) \right] \leq (n+1) \pi_\star \left(G_\star \left(\cdot, \varphi^p \right) \right) + C \nu_\star(V_\star) \; .$$

Combining these discussions imply that there exists a finite constant C_1 such that

$$\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} \{\varphi(Y_i) - G_\star(X_i, \varphi)\} \right|^p \leq C_1 \; n^{p/2 \vee 1} \; \nu_\star(V_\star) \; .$$

We now consider the second term in (63). Following the same lines as in the proof of (12, Proposition 12) and applying the Burkholder's inequality for martingales (13, Theorem 2.10), there exists a constant $C_2 < \infty$ such that

$$\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star} \left| \sum_{i=0}^{n} G_\star(X_i, \varphi) \right|^p \leq C_2 \; n^{p/2 \vee 1} \; \nu_\star(W_\star) \; .$$

The result follows.

PROOF. [Proof of Proposition 10] The first statement follows from standard results on phi-irreducible Markov chains satisfying the Foster-Lyapunov drift condition (20). By Lemma 18, for any $x \in \mathsf{X}$ and $F \in \mathcal{G}_{W_\star}$,

$$\mathbb{E}^{Q_\star}_x \left[\exp\left(\sum_{k=0}^{n} F(X_k) \right) \right] \leq V_\star(x) \mathrm{e}^{(n+1)(b_\star + \sup_{\mathsf{X}}(F - W_\star))} \; . \tag{64}$$

Since under the probability $\mathbb{P}^{T_\star}_{\nu_\star \otimes G_\star}$ the random variables $Y_{0:n}$ are conditionally independent given $X_{0:n}$, and the conditional distribution of Y_i given $X_{0:n}$ is $G_\star(X_i, \cdot)$,

$$\begin{aligned}\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star}\left[\prod_{k=0}^{n} \mathrm{e}^{\lambda_\star \varphi(Y_k)}\right] &= \mathbb{E}^{Q_\star}_{\nu_\star}\left[\mathbb{E}^{T_\star}\left\{\prod_{k=0}^{n} \mathrm{e}^{\lambda_\star \varphi(Y_k)} \middle| X_{0:n}\right\}\right] \\ &= \mathbb{E}^{Q_\star}_{\nu_\star}\left[\prod_{k=0}^{n} G_\star\left(X_k, \mathrm{e}^{\lambda_\star \varphi}\right)\right] \leq \mathbb{E}^{Q_\star}_{\nu_\star}\left[\exp\left(\sum_{k=0}^{n}\left|\log G_\star\left(X_k, \mathrm{e}^{\lambda_\star\varphi}\right)\right|\right)\right] .\end{aligned}$$

By the Jensen inequality, $F \stackrel{\text{def}}{=} \log G_\star\left(\cdot, \mathrm{e}^{\lambda_\star\varphi}\right)$ is non negative and belongs to $\mathcal{G}_{W_\star}$; we may thus apply (64) which yields

$$\mathbb{E}^{T_\star}_{\nu_\star \otimes G_\star}\left[\prod_{k=0}^{n} \mathrm{e}^{\lambda_\star \varphi(Y_k)}\right] \leq \nu_\star\left(V_\star\right) \mathrm{e}^{(n+1)(b_\star + \sup_{\mathsf{X}}(F - W_\star))} .$$

The proof then follows by applying the Markov inequality.

A Technical Results

We have collected in this section the proof of some of the technical results.

Lemma 17 *For any integer $n \geq 1$, and sequence $\mathbf{x} \stackrel{\text{def}}{=} \{x_i\}_{i \geq 0} \in \{0,1\}^{\mathbb{N}}$, denote by $M_n(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} \mathbb{1}\{x_i = 1\}$ and $N_n(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} \mathbb{1}\{x_i = 1, x_{i+1} = 1\}$. Then,*

$$M_n(\mathbf{x}) \leq \frac{n+1}{2} + \frac{N_n(\mathbf{x})}{2} .$$

PROOF. Denote by τ the shift operator on sequences defined, for any sequence $\mathbf{x} \stackrel{\text{def}}{=} \{x_i\}_{i \geq 1}$, by $[\tau \mathbf{x}]_k = x_{k+1}$. Let $\mathbf{x} = \{x_i\}_{i \geq 0}$ be a sequence such that $x_j = 0$ for $j \geq n$. By construction, $N_n(\mathbf{x}) = M_n(\mathbf{x} \,\mathrm{AND}\, \tau\mathbf{x})$. The proof then follows from the obvious identity:

$$\begin{aligned}n \geq M_n(\mathbf{x} \,\mathrm{OR}\, \tau\mathbf{x}) = M_n(\mathbf{x}) + M_n(\tau\mathbf{x}) - M_n(\mathbf{x} \,\mathrm{AND}\, \tau\mathbf{x}) \\ \geq 2M_n(\mathbf{x}) - 1 - N_n(\mathbf{x}) ,\end{aligned}$$

where AND and OR is the componentwise incluse ”AND” and ”OR”.

Lemma 18 *Assume that there exist a function $V : \mathsf{X} \to [1, \infty)$, a function $W : \mathsf{X} \to (0, \infty)$ and a constant $b < \infty$ such that*

$$\log(V^{-1}QV) \leq -W + b \; . \tag{A.1}$$

Let n be an integer and F_k, $k = 0, \ldots, n-1$, be functions belonging to $\mathcal{G}_W$, where $\mathcal{G}_W$ is defined in (26). Hence, for any $x \in \mathsf{X}$,

$$\mathbb{E}_x^Q \left[\exp\left(\sum_{k=0}^{n-1} |F_k(X_k)| \right) \right] \leq V(x) \mathrm{e}^{bn + \sum_{k=0}^{n-1} \sup_{\mathsf{X}} (|F_k| - W)} \; . \tag{A.2}$$

PROOF. The proof is adapted from (17, Theorem 2.1). Set for any integer n,

$$M_n \stackrel{\text{def}}{=} V(X_n) \exp\left(\sum_{k=0}^{n-1} \{W(X_k) - b\} \right) \; . \tag{A.3}$$

The multiplicative drift condition (A.1) implies that $\{M_n\}$ is a supermartingale. Hence, for any $n \in \mathbb{N}$ and $x \in \mathsf{X}$,

$$\mathbb{E}_x^Q \left[V(X_n) \exp\left(-bn + \sum_{k=0}^{n-1} W(X_k) \right) \right] \leq V(x) \; .$$

The proof follows.

References

[1] C. Andrieu and A. Doucet. Particle filtering for partially observed Gaussian state space models. *J. Roy. Statist. Soc. Ser. B*, 64(4):827–836, 2002.

[2] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(6):697–725, 1997.

[3] A. Budhiraja and D. Ocone. Exponential stability of discrete-time filters for bounded observation noise. *Systems Control Lett.*, 30:185–193, 1997.

[4] A. Budhiraja and D. Ocone. Exponential stability in discrete-time filtering for non-ergodic signals. *Stochastic Process. Appl.*, 82(2):245–257, 1999.

[5] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

[6] P. Chigansky and R. Lipster. Stability of nonlinear filters in nonmixing case. *Ann. Appl. Probab.*, 14(4):2038–2056, 2004.

[7] J. Dedecker and P. Doukhan. A new covariance inequality and applications. *Stochastic Process. Appl.*, 106(1):63–80, 2003.

[8] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[9] P. Del Moral and A. Guionnet. Large deviations for interacting particle systems: applications to non-linear filtering. *Stoch. Proc. App.*, 78:69–95, 1998.

[10] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré*, 37:155–194, 2001.

[11] P. Del Moral, M. Ledoux, and L. Miclo. On contraction properties of Markov kernels. *Probab. Theory Related Fields*, 126(3):395–420, 2003.

[12] G. Fort and E. Moulines. Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.*, 31(4):1220–1259, 2003.

[13] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, New York, London, 1980.

[14] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *J. Finance*, 42:281–300, 1987.

[15] E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models (with discussion). *J. Bus. Econom. Statist.*, 12:371–417, 1994.

[16] M.L. Kleptsyna and A.Y. Veretennikov. On discrete time ergodic filters with wrong initial conditions. Technical report, Université du Maine and University of Leeds, 2007. Available at http://www.univ-lemans.fr/sciences/statist/download/Kleptsyna/filt19fg.pdf.

[17] I. Kontoyiannis and S. P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.*, 10:no. 3, 61–123 (electronic), 2005.

[18] François LeGland and Nadia Oudjane. A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Process. Appl.*, 106(2):279–316, 2003.

[19] A. Manrique and N. Shephard. Likelihood inference for limited dependent processes. *Econometrics Journal*, 1:174–202, 1998.

[20] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, London, 1993.

[21] D. Ocone and E. Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control*, 34:226–243, 1996.

[22] Nadia Oudjane and Sylvain Rubenthaler. Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.*, 23(3):421–448, 2005.